

東芝クラウドソーシングの構築と活用

著者	芦川 将之, 川村 隆浩, 大須賀 昭彦
雑誌名	情報処理学会デジタルプラクティス
巻	9
号	4
ページ	823-841
発行年	2018-10-15
URL	http://id.nii.ac.jp/1438/00009004/

特集招待論文

東芝クラウドソーシングの構築と活用

芦川 将之¹ 川村 隆浩² 大須賀 昭彦²

¹東芝デジタルソリューションズ（株） ²電気通信大学大学院情報システム学研究科

クラウドソーシングはCrowd(群衆)+Sourcing(調達)の造語であり、「企業、組織が、自社もしくはアウトソースの人材により実施していた業務を、よりオープンかつ不特定多数のCrowd(群衆)から人材を集め実施すること」と定義されている。我々は大量の研究データの解析を低コストかつ高速に処理するためにマイクロタスク方のクラウドソーシングの利用を試みている。しかし従来のクラウドソーシングサービスでは作業結果を研究に利用するには精度の点で問題があり、さらに精度を向上させるために外部のサービスが提供している機能の範囲では十分ではなく、外部のサービスに新規の機能を追加することも難しいという問題がある。我々はこれらの問題を解決するために、独自の東芝クラウドソーシングシステムを構築し、システム内にて精度向上手法を適用することで問題の解決を試みている。本稿では東芝クラウドソーシングにおける精度向上手法、及び活用事例に関して紹介する。

1. はじめに

クラウドソーシングはCrowd（群衆）+Sourcing（調達）の造語であり、「企業、組織が、自社もしくはアウトソースの人材により実施していた業務を、よりオープンかつ不特定多数のCrowd（群衆）から人材を集め実施すること」と定義されている。企業などが目的（需要）を提示し、それを不特定多数の情報発信者が参加して解決（供給）することで大量の作業を効率よく処理することが目的である。従来は不特定多数の人間に対して目的を提供、結果の収集を行うことが難しかったが、インターネットの技術革新に伴い可能となった。

大学・企業等の研究機関では、このクラウドソーシングの技術をさまざまな研究データの解析に用いている。研究データの作成は精度的な問題から自動化できないケースが多く、研究者、もしくは専門の技術を持った外部の業者といった人手による作業が必要になる。しかし、昨今の研究に用いられるデータはビッグデータと称される大量なデータであることが多い。従来の人手による作業では巨大データを扱うにはコスト、速度の面から難しくなっている。

既存のクラウドソーシングサービスとして、数多くのサービスが存在する。しかし、これらの外部サービスを研究データの作成に利用するにはデータの機密性の保持と精度の面から問題があった。企業が保持する研究データは秘匿性が高いデータが多く、外部のサービスに委託するには機密上、さまざまな点で問題が発生する。このような企業内のデータの機密性を保持するためには外部のサービス利用では難しい。さらに、データを利用する目的は、自然言語処理における言

語モデルの構築など、得られたデータから統計的なデータを用いるため精度の重要性が低いケースから、実験結果データの評価、自然言語処理における辞書データの構築など精度の重要性が高いケースまでさまざまなケースがある。このような精度の重要性が高いケースでは作業結果の品質を高く維持しなくてはならないが、そのためには外部のサービスが提供している機能の範囲では十分ではなく、さらに外部のサービスに新規の機能を追加することも難しい。我々はこれらの問題を解決するために、クラウドソーシングシステムを機密性が高くデータの安全性を高めることが可能なプライベートな環境下において構築することで問題の解決を試みている。

本稿では我々が構築、活用している東芝クラウドソーシングシステムを紹介する。

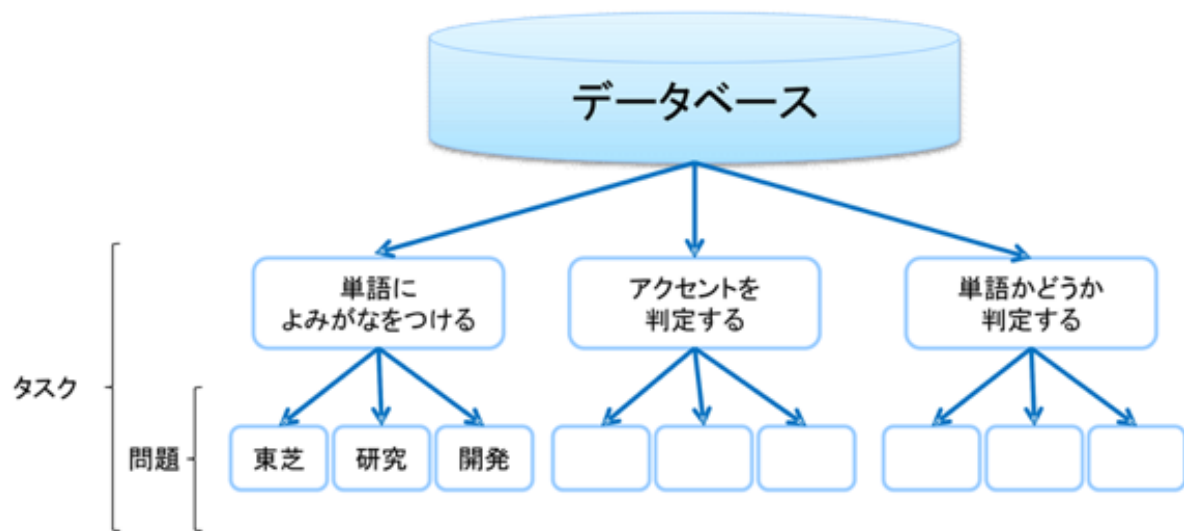
2. 東芝クラウドソーシング

研究データの構築には大量のタスクを高速に処理しなければならない。そのため、我々は大量のタスクを数多くの不特定の作業者に処理させるマイクロタスク型のクラウドソーシングを用いている。しかし、既存のマイクロタスク型のクラウドソーシングサービスを研究データ構築に利用するには精度の点に問題がある。

そこで、これらの問題を解決するために、プライベートな環境下においてさまざまな精度向上手法を適用したマイクロタスク型のクラウドソーシングシステムである東芝クラウドソーシングを構築した[1]。

2.1 東芝クラウドソーシング上で処理される作業の分類

クラウドソーシング上で処理するさまざまな作業はタスクと呼称され、さまざまなタスクが存在する。規模も研究テーマの考案のような大きなタスクから、アンケートなどの小さなタスクまで多岐に渡る。本研究ではマイクロタスク型のクラウドソーシングを対象としているため、処理が数秒から数分で完了するような小規模な作業が主な対象となる。しかし、作業のサイズが小さくなると個々の作業を管理するのは煩雑になるため、クラウドソーシングでは同様の小さな作業をまとめて処理することが多い。東芝クラウドソーシングではこのまとまりを「タスク」と呼称している。たとえば図1における「単語に読み仮名をつける」作業を東芝クラウドソーシングで行う場合、1つひとつの単語に読み仮名をつける作業を「問題」、「1,000 問の単語に読み仮名をつける」という作業の集合が「タスク」となる。出題者はこのタスク単位で東芝クラウドソーシングに作業を出題する。東芝クラウドソーシングではこの出題者を「リクエスト」と呼称している。



- ・複数の同一種類の問題をまとめたものが「タスク」
- ・出題者(リクエスタ)はタスク単位で出題を行う

図1 東芝クラウドソーシングにおけるタスク

東芝クラウドソーシングの主な利用用途としてはデータの作成であることは述べたが、大きく分けてデータの作成には何もないところからテーマやルールに従ってデータを作成する「データ収集・作成」系とすでに存在するデータをベースに精錬化，別系統のデータへの変更などを行う「データ加工」系が存在する（図2）。また，「データ収集・作成」系で作成したデータをさらに「データ加工」系のタスクで処理するケースも存在する。これらのタスクに関する情報は東芝クラウドソーシング内のリクエスタ間で共有されており，既存のタスク作成やタスクシナリオにおけるノウハウを共有することで経験が少ないリクエスタも初回から精度の高い結果を得ることができている。

データ収集・作成		データ加工	
データ評価	データ収集	データ付与	データ変換
アンケート	例文作成	単語読み入力	画像処理
品質評価	音声収集	単語品詞入力	音声処理
内容判定	テーマ提案	単語アクセント付与	言い換え表現
文の自然性判定			略称作成
文章判定			

図2 東芝クラウドソーシングにおけるタスク例

東芝クラウドソーシングシステムは図3のような構成となっている。リクエスタはWebインタフェース経由でタスクをデータベースに登録し，ワーカはデータベースに登録されたタスクに対してWebインタフェース経由でタスク処理を行い，結果をデータベースに登録する。リクエスタ

はタスク処理が完了次第、結果をデータベースから取得する。次節ではこれらの流れをワーカ視点で説明する。

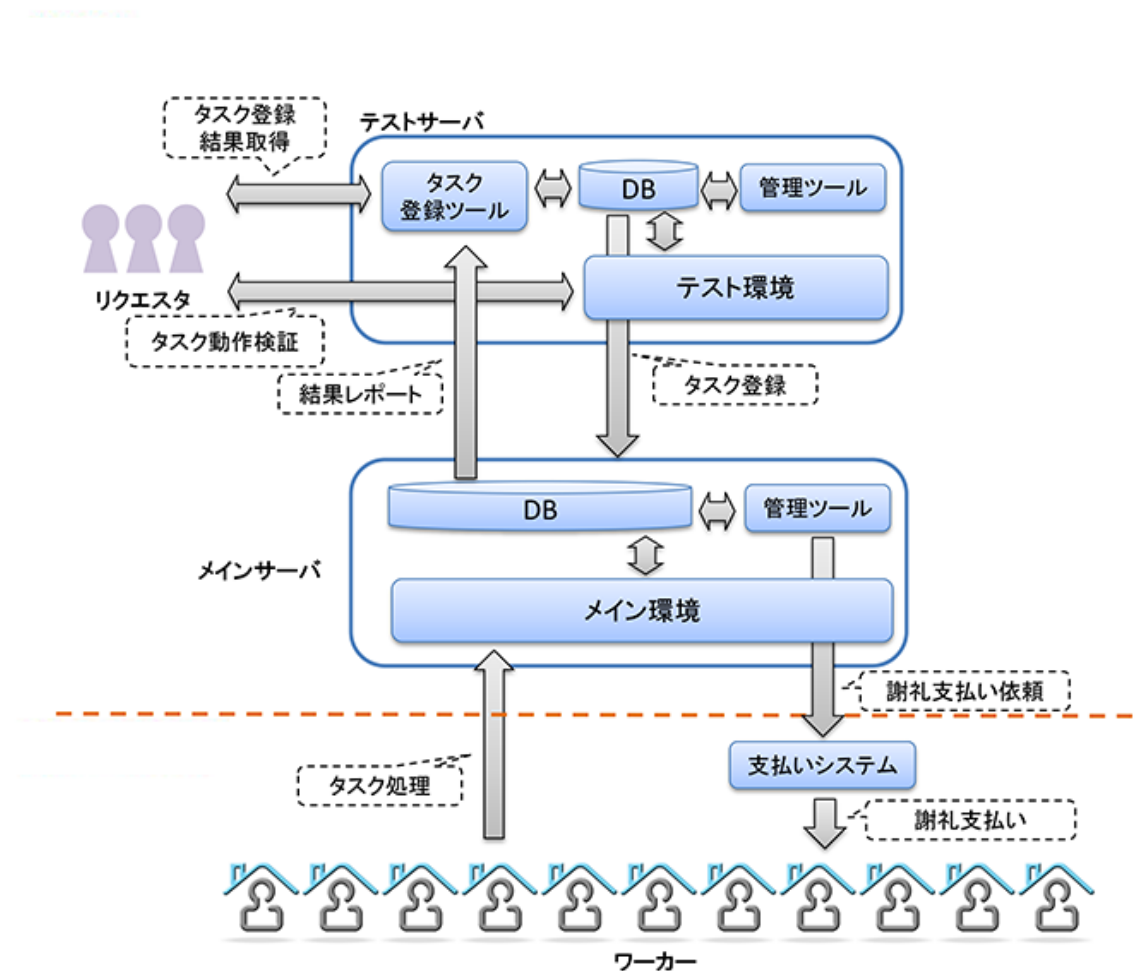


図3 東芝クラウドソーシングのシステム構成

2.2 ワーカ視点での東芝クラウドソーシングにおける処理の流れ

図3の構成はワーカ視点では図4のようになる。東芝クラウドソーシングに登録されたワーカは東芝クラウドソーシングのシステムにログインを行う。ワーカはログイン後、1) タスク選択フェーズ、2) トレーニングフェーズ、3) タスク処理フェーズを繰り返して作業を進めていく。実際のタスクはリクエスタがデザインしているため、さまざまな入力インタフェースが存在する。ワーカは自分のステータスを確認しながら作業を進めていく。タスク処理フェーズの画面例を図5に示す。作業者は画面左の作業説明画面と結果入力画面に対して作業を進めていく。画面右には作業者のステータスが常時表示されており、作業者は自分のステータスを確認しながら作業を進めていくことができる。精度が一定以下になると作業ができなくなることを作業者に通知しているため、作業者は自分の結果精度が下がらないよう気をつけて作業を進め、結果として精度向上へつなげることができる。

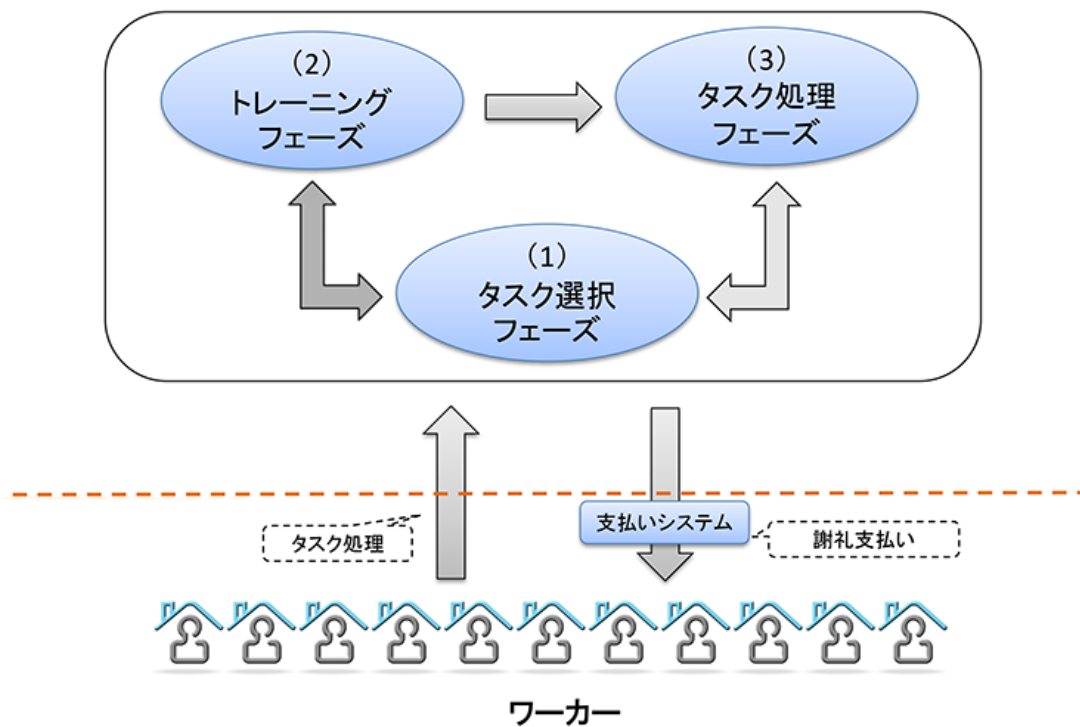


図4 ワーカ視点での東芝クラウドソーシングにおける処理の流れ

クラウドソーシング - 日本語研究のための簡単なデータ入力作業 -

>> 作業一覧

作業 単語の読みを入力する (66)

表示された言葉の読み仮名を入力してください。

問題: 落ち枝 残り時間: 98秒

読み (ひらがな):

回答

[落ち枝をGoogleで検索]

課題の異常を報告する 回答せずに次の作業へ 作業を終了する

あなたのステータス

総作業数:	504件
確認済み作業数:	203件
確認中作業数:	300件
今月獲得マイル:	643mile
(上限は60,000mile)	
先月獲得マイル:	164mile
(12/28から1/30まで)	
正解率:	98.5%
経験値:	197exp

図5 タスク処理フェーズ

2.3 東芝クラウドソーシングのワーカ

東芝クラウドソーシングで作業を行うワーカは複数のグループで管理されており、その特徴に応じて適したタスクに割り当てられる。主なワーカのグループは以下の4グループとなる。

(1) プライベートクラウドワーカ

クラウドソーシングの対象候補となった作業が実際に処理可能かどうか、生産性や品質を検証するために作業する実験用のワーカグループ、主に高セキュリティ環境下で作業することが可能であるがグループとしては小規模である

(2) デディケイテッドワーカ

主に高難易度のタスクや品質チェックを行う高スキル、高品質のワーカグループ、主に高セキュリティ環境下で作業を実施するため秘密保持性も高いが、コストが高いためその他のグループと組み合わせて作業を行うケースが多い。

(3) コミュニティクラウドワーカ

東芝における契約ワーカ、デディケイテッドワーカと比較して低コストで作業を実施することが可能である。ネットワーク経由の個人環境から作業を行うが全員NDAを締結して作業を行うため秘密保持性も比較的高い。

(4) パブリッククラウドワーカ

社外の人材から募集しているワーカ、全体人数が多いため生産性が高く、デディケイテッドワーカと比較して低コストだが秘密保持性が低い。

東芝クラウドソーシングでは上記のワーカグループを用途に応じて組み合わせて作業を実施している。たとえばセキュリティ要求が高い作業において、秘密保持性の高いデディケイテッドワーカでセキュリティ対象個所の選別、処理を行いコミュニティワーカでセキュリティ要求の低い個所の処理を行うことでコスト削減を実現するなどである。

3. ワーカのフィルタリングによる精度向上手法の提案

クラウドソーシングの環境を構築しただけでは得られるタスクの処理結果の精度は低いため、研究データとして使用するには十分ではない。本章では東芝クラウドソーシングにおけるワーカフィルタリングによるタスク処理結果の精度向上手法に関して述べる。

3.1 ワーカを対象とした精度向上手法

東芝クラウドソーシングにおける精度向上手法は主にワーカに対する管理を中心に行っている。クラウドソーシングは「不特定多数の外部の人間」に作業を委託する仕組みであるため、ワーカの品質はさまざまである。特定のカテゴリのタスクにおける正解率が高い高品質ワーカや正解率が低い低品質ワーカ、リクエストの出題意図に沿った回答ができるスキル保持ワーカや意図に反した回答をする負スキル保持ワーカ、全体の正解率が低い、またはスクリプトなどで処理を行う、システムから排除対象となるスパムワーカなどのワーカが存在する。既存のクラウドソーシングサービスでは数多くのリクエストから数多くのタスクを受け入れているため、ワーカが行うタスクは多種多様となり、結果としてタスク単位におけるワーカの行動情報が少なくなり、ワーカのコントロールが難しくなっている。東芝クラウドソーシングではプライベートという特徴上タスクのカテゴリが限られているため、タスクカテゴリに対するワーカの行動情報は相対的に多くなっており、そのワーカの行動情報を活かすことでワーカの特性に合った適切なタスクを与え、低品質ワーカおよびスパムワーカを排除することを可能としている。

以下にワーカに対する東芝クラウドソーシングの精度向上手法を（1）事前フィルタリング、（2）動的フィルタリング、（3）結果フィルタリング、（4）推測フィルタリング、の4つのカテゴリに分類した。それぞれのフィルタリングではコストと精度が異なり、コストが高いフィル

タリングは低品質ワーカを排除する精度が高い。我々はコストの問題から、対象のワーカの数に応じてフィルタリングを適用している。それぞれの手法は東芝クラウドソーシングの運用における図6に示したタイミングで行われる。それぞれのフィルタリングに関して詳細を述べる。

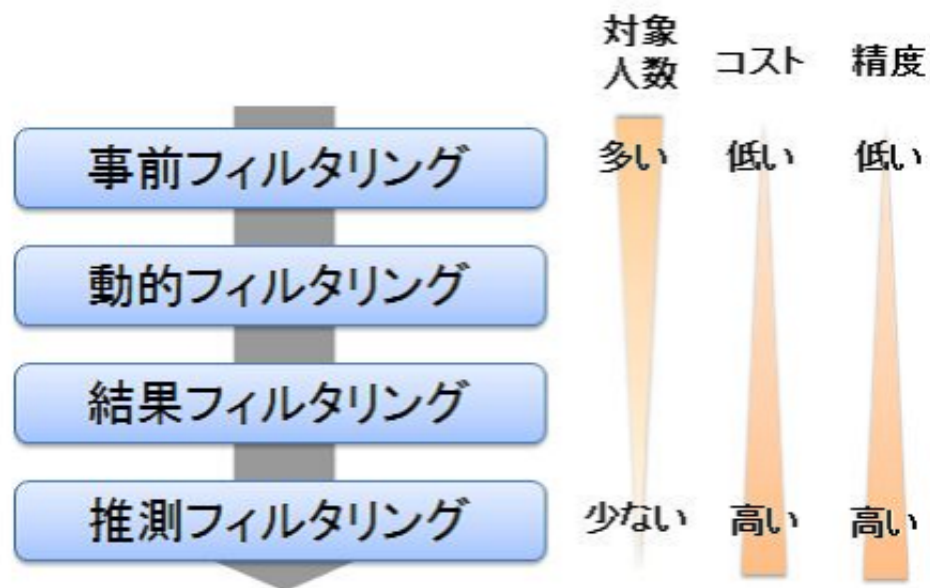


図6 東芝クラウドソーシングにおけるワーカに対する精度向上手法

3.2 事前フィルタリング

ワーカを募集する際に行うフィルタリングである。東芝クラウドソーシングシステムの作業者はOB・OGを含めた関係者や外部の人材など用途に応じて多岐にわたり、作業候補となる人材も数多い。しかし、これらのすべての会員をワーカとして扱うのはコスト的に現実的ではなく処理能力的にも過剰である。また、ワーカ候補の素養もさまざまであるため、適したワーカを抽出するために事前のアンケートを用いてフィルタリングを実施するなどしている。また、難易度の高いタスクでは、対象となるタスクを処理できるワーカに特化して募集するために問合せ内容に追加して、対象となるタスクの出題内容を一部出題することでフィルタリングを行う場合もある（図7）。

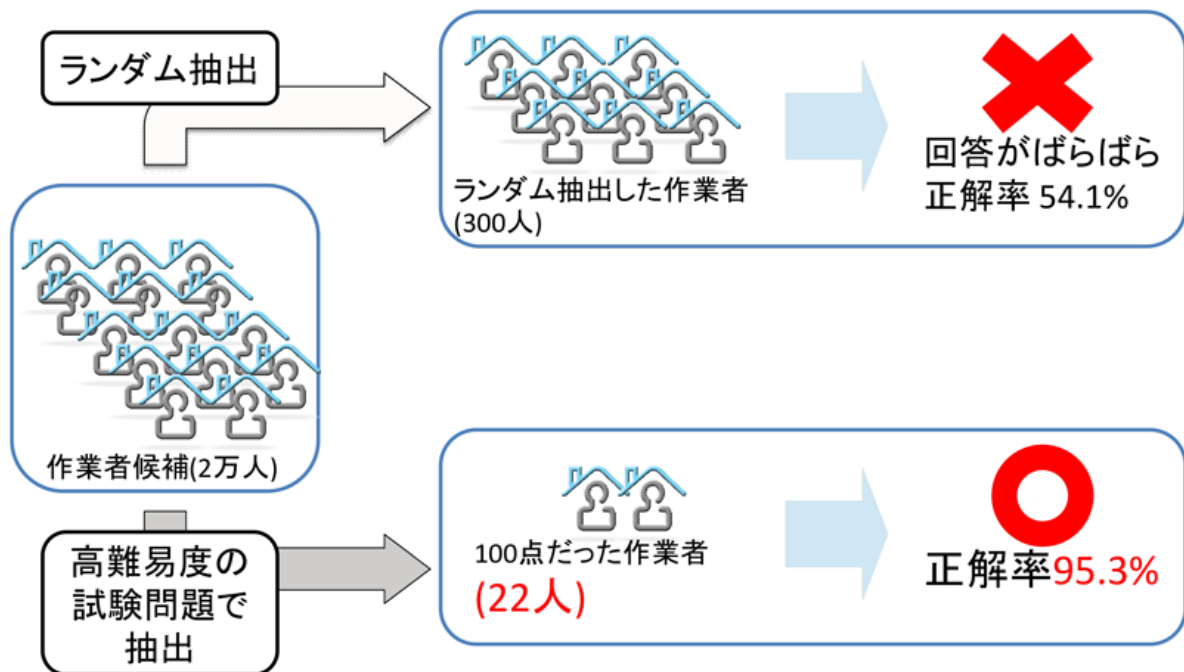


図7 事前フィルタリング

3.3 動的フィルタリング

ワーカがタスク処理をしている際に行うフィルタリングである。事前フィルタリングにて最低限の品質を確保できたワーカであるが、すべての低品質なワーカを排除できたわけではない。また、人間は時間の経過に応じて能力が上下するため、初期の品質判定が継続するとは限らない。そのため、タスク処理を進めていく過程で動的にワーカのフィルタリングを行うために正解率と経験値という2点の項目を設けている。

正解率は「正解数／総作業数」で算出し、一定値以下のワーカは低品質ワーカと見なし、以降の東芝クラウドソーシングにおけるタスク処理を禁止する。正解率を求めるにあたって、それぞれの作業はあらかじめ正解が用意されていないため、ワーカの処理結果が正解であるか否かを判定する手法が必要となる。この合否判定に用いる手段としては多数決を用いる手法が提案されている[2]。我々も主に多数決にて正解を決定しており、アンケートなど正解がない場合にはタスク説明に正解がない旨を明記し、正解率は変動させない。多数決が用いられるタスクは非常に多岐に渡るが、例として選択式のタスクが挙げられる。実際にこの作業をワーカに処理させたところ、回答において3人が一致した率は79.4%、2人が一致した率は19.0%、バラバラだった率は1.3%、未回答は0.3%となった。そして3人が一致したケースに対して、作業結果から1,000件をランダムで抽出、システム管理者側で手動で正解を作成して、正解率をチェックしたところ95.3%と非常に高い正解率を得ることができた。そのため、一般的なタスクの出題方法として、3人一致したケースのみデータとして採用し、残りのデータは不採用、もしくは再度クラウドソーシングに出題するというパターンを用いている。

正解率が一定値以下になることでタスク処理ができなくなることはワーカに明言しており、ワーカはこの数値表示によって精度に対する注意を喚起される。これらの数値は作業員の行動によって変化するという点でゲームメカニクスにおける得点制度と同等に考えることができる。得点制度はゲームメカニクスとしては一般的であり、利用者のモチベーションを向上させるための手法として用いられている[3]。

また、同様に、「正解数-不正解数」で算出される経験値を設定し、一定の経験値を持つワーカに対して高報酬、高難易度のタスクを提供している。難易度の基準はリクエストによって異なるが、多数決による正解判定を行うタスクで結果が分散してしまう、タスクの完了まで時間がかかるなどのケースでは難易度が高いと判定される場合が多い。これらの数値は図8のように作業中に画面に常に表示している。



図8 ワーカに表示されるステータス画面

現時点ではインタフェース上の制限から、ワーカが確認することができるのはすべてのタスクの全体平均正解率のみである。しかし、動的フィルタリングをこの全体平均正解率のみで行うとフィルタリング効果が低いことがわかっている。我々は表1のようにカテゴリごとに正解率をワーカに明示せず別途管理している。本研究では第4章で事例として紹介した自然言語処理に関する「単語判定カテゴリ」「読みつけカテゴリ」「品詞カテゴリ」「アクセントカテゴリ」に関して述べる。表1を見ると正解率が低いカテゴリの作業総数が非常に少なく、正解率が高いカテゴリの作業総数が多いケースが散見される。これにより正解率が下がるような難易度の高いタスクをワーカが避ける傾向にあることが分かる。しかし、全体正解率のみで判断を行った場合、「賃金が高く難易度も高いタスクA」と「賃金が低く難易度も低いタスクB」があった場合、ワーカはタスクAを処理し、全体平均正解率が下がるとタスクBを行って全体平均正解率を回復させるという行動をとることがあった。これは該当するカテゴリの正解率が低いにもかかわらず大量に作業を処理しており、かつ全体正解率が高いというワーカの存在から判明した。これらのワーカを低品質ワーカと呼称している。このような低品質ワーカの行動に対応するため特定カテゴリにおける作業総数が50以上になったワーカをアクティブワーカとし、特定のカテゴリにおけるアクティブワーカの精度が一定値以下になった場合は、そのカテゴリに属するタスクを隠し、処理をさせないようにすることでワーカの行動コントロールを行っている。東芝クラウドソーシングでは該当カテゴリのタスクにおける精度が60%以下になったワーカはその作業をさせないという対

応をしている。これは70%以下のワーカを排除対象とすると悪意のないワーカを多く排除してしまい、50%以下のワーカを対象とすると、悪意のあるワーカを排除しきれなかったというシステム運用上の経験からの数値である。

表1 タスク別ワーカ結果精度（一部）

ワーカ-ID	読み付け			読み仮名判定			品詞判定		
	正解率	正解位数	不正解数	正解率	正解位数	不正解数	正解率	正解位数	不正解数
101	92.6	14368	1147	94.6	4295	244	88.6	474	61
102	96.2	53463	2086	97.1	23028	695	83.6	504	99
103	97.1	1455	43	98.3	5385	94	100.0	10	0
104	94.5	10247	597	95.8	28824	1250	83.9	3406	654
105	91.9	452	40	0.0	0	0	0.0	0	0
106	98.0	16010	329	99.3	11558	82	93.2	68	5
107	95.2	64775	3240	94.7	42631	2398	88.7	11915	1517
108	97.0	44965	1375	98.2	39815	741	89.4	1831	216
109	97.4	69290	1863	97.9	25541	543	84.4	1094	202
110	96.4	65581	2462	97.0	29294	903	90.5	7435	780
111	94.6	2164	124	95.7	44	2	100.0	13	0
112	90.4	64183	6792	95.0	52971	2814	82.4	17895	3833
113	95.9	77179	3313	96.1	94042	3841	89.8	13512	1532
114	95.5	121979	5746	95.9	85658	3629	47.1	3985	4483
115	90.0	98895	11040	97.0	114462	3533	92.7	11622	912

3.4 結果フィルタリング

図9のように、ワーカのタスク処理結果からワーカの特徴を判別するフィルタリングである。動的フィルタリングは正解を判定することができる作業に対してのみ有効であり、アンケートや文章作成のような明確な正解がなく、多数決も実施しにくいタスクにおいては適用できない。また、ワーカが低品質ワーカであると判明し、出題停止に至るまでに多くの低品質なデータが算出されてしまうという欠点がある。我々はこの問題に対し、図9のように、ワーカのタスク処理結果からワーカの特徴を判別する結果フィルタリングを用いて対応している。

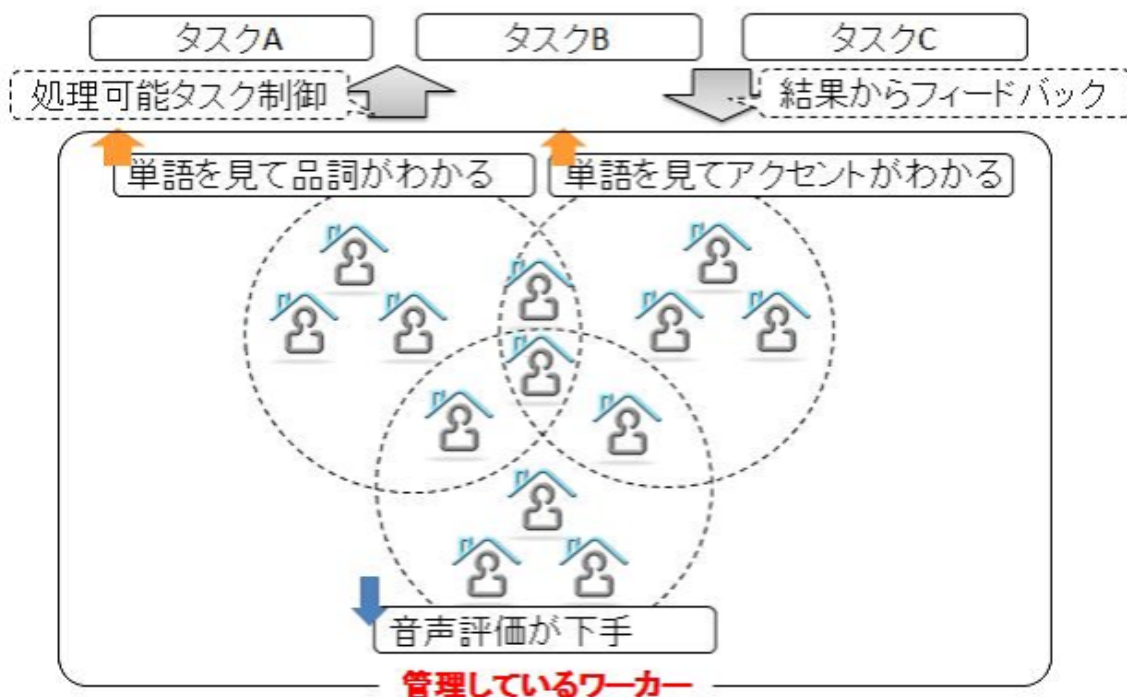


図9 タスク処理結果を用いたワーカーの特徴付け

明確な正解がないタスクでも、リクエストの意図に沿った内容か否かという判定は存在しており、この判定をリクエストにタスクごとに行わせるには大きなコストがかかる。このようなタスクに関して、リクエストはほかのリクエストの類似したタスクの結果や、小規模のテスト用タスクを実施した結果などから、正解率の高いワーカーや出題意図に沿った回答をしているワーカーを選別し、以降のタスクは条件に該当するワーカーのみに出題することで結果精度を向上させることができる。この選別基準はシステム側で明確に定めておらず、リクエストによって異なる。これらのワーカーの情報で優秀なワーカーを判別するための情報を「スキル」、低品質なワーカーを判別するための情報を「負スキル」と呼称している。「負スキル」はカテゴリごとに作成可能であり、「負スキル」保持ワーカーは該当するカテゴリのタスク以外は作業できるため、すべての作業が不可能となっているスパムワーカーとは異なる。たとえば「品詞」のカテゴリのタスクの正解率が高いワーカーには「品詞」のスキルを付与し、「品詞」のタスクは「品詞」スキルを持つワーカーのみ出題することで精度向上を行っている。これらのスキルはリクエスト間で共有して使用することができるため、新規のリクエストも初回からスキルを保持するワーカーにタスクを処理させることが可能である。

3.5 推測フィルタリング

動的フィルタリングや結果フィルタリングは何らかのタスクの処理結果をワーカーの行動コントロールに流用したものであり、ワーカーがスパムワーカー、低品質ワーカーであった場合はワーカーの行動コントロールができる段階に達した時点で低品質な処理結果を残してしまっていることが多い。これらのデータは再処理が必要であり、大量のワーカーによって短時間で大量のタスク処理が行われるマイクロタスク型のクラウドソーシングでは時間、賃金ともに再処理のコストが大きくなってしまふ。そこで、我々はさらに低品質なタスク処理結果を削減するために、ワーカーの特性から行動を推測し、事前にタスクに不適切なワーカーをフィルタリングする試みを行っている（図10）。

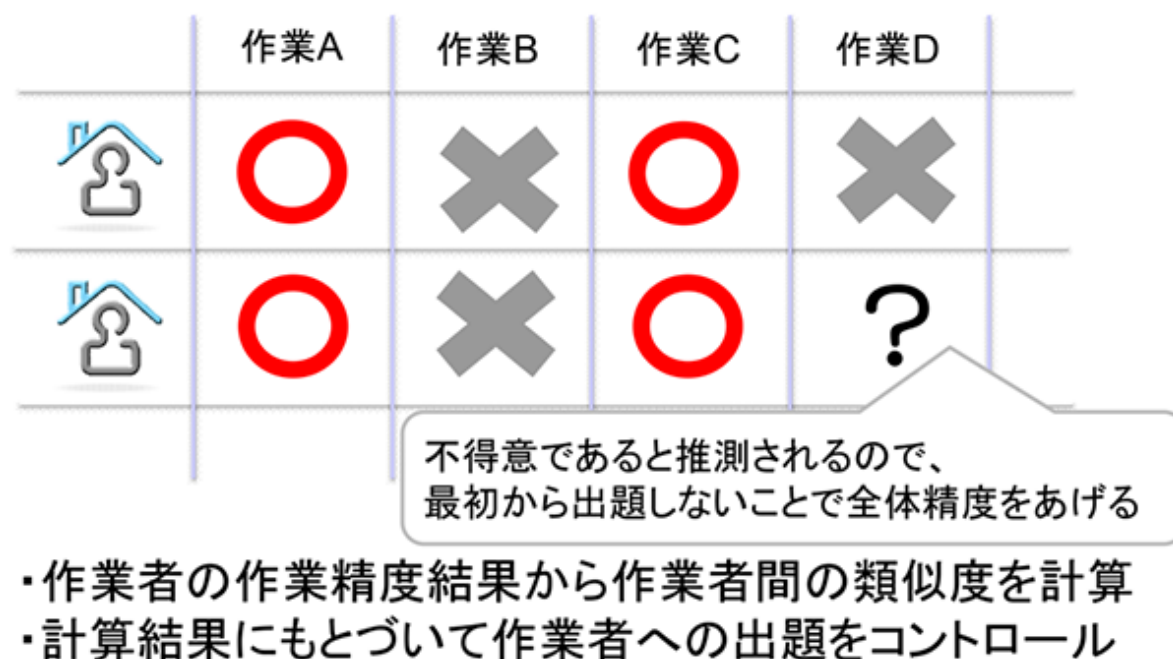


図10 推測フィルタリング

このようなワーカに対するタスクの割り当てに関する研究としてさまざまな研究がなされている。タスクの内容やワーカのタスクに対する完遂率をベースにタスクの推薦を行う研究[4]では低品質ワーカに対する対応が取られておらずタスク推薦の効果があらわれるまでに多くの低品質データが発生してしまう問題がある。我々は推測フィルタリングに至るまでの複数のフィルタリングで低品質ワーカを可能な限り少なくすることで、低品質データの発生を最低限におさえている。また、ワーカの行動履歴、ワーカのタスクに対する嗜好からワーカにタスクの推薦を行う研究[5]でも対象となるワーカが膨大になった場合のコストが大きいという問題がある。我々は前述のように推測フィルタリングに至るまでの複数のフィルタリングで対象となるワーカの数削減し、必要なコストを最低限に抑えている。また、タスクの難易度レベル、ワーカのスキルのレベルを推測した結果からワーカにタスクの推薦を行う研究[6]でも対象となるタスクのカテゴリが限られているという問題がある。我々は複数のカテゴリを管理し、タスクをカテゴリに分類することで複数のタスクカテゴリを対象とすることを可能としている。

4. 東芝クラウドソーシングの活用事例

東芝クラウドソーシングを用いて知識処理研究に必要な語彙を収集した事例について述べる。まず始めに、Webクローラを用いた大規模テキストの収集を行い、続いて収集したテキストから未知語の候補を自動抽出する。そして最後に、東芝クラウドソーシングを用いて未知語候補から単語として適当なものだけを絞り込み、知識処理研究の一環である音声認識や音声合成の辞書を構築するために必要な品詞や読み仮名、アクセント等の単語情報を付与する。

4.1 語彙の重要性

知識処理研究では基礎となるデータとして大量の語彙情報を必要とする。たとえば現在の形態素解析では辞書を用いるが、その辞書で語彙が不足していた場合、未知語が多く発生してしまううまく形態素解析を行うことができず、望む結果が得られないことがある。たとえば形態素解析を用いる研究の例として電子書籍読み上げが挙げられる未知語による読み誤りは精度を低下させる大きな原因となっている。また、新語は常に発生しており、それらの新語を語彙として常に辞書に登録する必要がある。そのためにはWebなどにおける最新の大量のテキストデータから語彙を抽出し、音声処理や自然言語処理といった知識処理研究に必要な情報である「読み仮名」「アクセント」「品詞」といった情報を付加していかなければならない。

4.2 未知語候補の抽出

クローリングでWebから収集したテキストから未知語の候補を抽出する。抽出処理は以下のステップで行った。

- (1) テキストに対して点予測手法[7]による単語分割を実施
- (2) 単語分割結果から辞書未登録文字列を取得
- (3) 単語分割結果を用いて単語Ngramを作成
- (4) 単語Ngramを用いて辞書未登録文字列の中から未知語候補を選出

この一連の処理によって125億文のテキストから23万語の未知語候補を抽出することができた。


4.3 単語判定と単語情報付与

4.2節の方法で作成された未知語候補には、単語として適当でないものが残っている可能性が高い。また、抽出した単語に対して音声処理に必要な情報を付与しなくてはならない。これらの情報収集を東芝クラウドソーシングの以下の4タスクとして行った。

(1) 単語判定タスク

タスクデザインを図11に示す。このタスクではワーカに対して4.2節の方法で作成された未知語候補を「それは（未知語候補）です」という問題文に加工して表示し、「問題文は日本語として自然か否か」という選択をさせた。「日本語として自然である」と回答された場合、その文章に含まれる未知語候補を未知語として扱う。たとえば図の例では抽出された語彙「里山」を単語かどうか判定するために「それは里山です」という例文を用いた。

「里山」は単語として判断されるのが理想であるため、この文章は問題であると回答されるのが望ましい。しかし、形態素解析の結果によっては「お子ちゃまと一緒に」という文から「ちゃま」という単語が未知語として抽出されてしまう場合がある。この場合は「それはちゃまです」という文が例文として提示される。「ちゃま」は単語として判断されないのが理想であるため、「ない」という結果が得られるのが望ましい。

 作業

例文において赤字で表示された文字列を
使用例のように使うことがあるかどうか判定してください。

残り時間：180秒

例文 : 薪棚、**里山**と田畑
 ↓
使用例 : **それは里山です。**

回答 : ☐ある
 ☐ない

**問題にアダルト、不快な内容が含まれる場合は
以下のボタンを押してください。**

図11 単語判定タスク

(2) 品詞付与タスク

タスクデザインを図12に示す。このタスクでは名詞とそれ以外の品詞に分ける作業を行っている。名詞に関しては「人名」「地名」「組織名」「その他の名詞」に再分類している。(1)で単語として適切であると判定された未知語に単語抽出元の前後の文章を付与して問題文に加工して表示し、「人名」「地名」「組織名」「その他の名詞」「名詞以外」を選択させた。



作業 単語の品詞を選択する

例文において赤字で表示された単語の品詞として最も適切なものを下の選択肢から1つ選んでください

残り時間：122秒

例文： キャップをボディエンドにつけ

- 回答：
- ☐ 人名
 - ☐ 地名
 - ☐ 組織名（会社名や団体・グループ名など）
 - ☐ その他の名詞
 - ☐ 名詞以外

回答

回答せずに次の作業へ

[[ボディエンドをGoogleで検索](#)]

課題の異常を報告する

回答せずに次の作業へ

作業を終了する

図12 品詞付与タスク

(3) 読み付与タスク

タスクデザインを図13に示す。このタスクでは(2)で名詞と判定された未知語を問題として表示し、その読みを入力させ、その結果を未知語に対する読みと判定した。最初は単語抽出を行わず文章への読みつけ作業を行った。1つの文章を3人に対して出題し、3人、もしくは2人が一致したものを正解とした。しかし、この結果としては3人一致が17.8%、2人一致が38.5%、不一致が37.3%とずれが大きい結果となった。これは長文入力において入力ミスなどの誤差が多く、長文のためミスの影響範囲が大きいなどが原因であると判断し、図13のように単語への読みつけを行う方式へと変更した。これにより精度が大幅に向上した。また、単語にしたということで一作業あたりの報酬を下げることで全体のコストを上げることなく精度改善を可能とした。

 作業

単語の読みを入力する(65)

表示された言葉の読み仮名を入力してください。

残り時間：98秒

問題：ブリヴェ企業再生グループ株式会社

読み（ひらがな）：

[\[ブリヴェ企業再生グループ株式会社をGoogleで検索\]](#)

図13 読み付与タスク

(4) アクセント付与タスク

タスクデザインを図14に示す。このタスクでは(3)で付けられた読みから推定されるアクセント候補から合成した音声を用い、どれが自然かを選択させた。その結果を未知語に対するアクセントと判定した。この作業は難易度が高いため、アクセントスキル保持ワーカ163名にのみタスク処理させている。



作業 言葉の共通語アクセントを選ぶ

NHKアナウンサーの発音を選択してください。

残り時間：174秒

問題：「無損傷 が （ムソンショウガ）」

全て再生

発音：

☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐

ム	ソ	ン	シ	ョ	ウ	ガ
---	---	---	---	---	---	---



☐ いずれでもない
読みが間違っている

回答

課題の異常を報告する

回答せずに次の作業へ

作業を終了する

図14 アクセント付与タスク

各タスクは3人に出題され、2人以上一致した回答を有効なデータとして扱う。ただし、(1)の単語判定タスクは高精度であることを求められるため、3人が一致した回答のみを有効なデータとして扱った。また、ワーカが設問が不適切であると判断した場合は「パス」を選択できるようにしている。通常のパスであれば回答権はほかのワーカに移動するが、6回以上パスが行われた場合はその問題は不適切と判定されて排除される。東芝クラウドソーシングではリクエストからの中断依頼がない限り、出題したすべての問題に対して回答かパスの処理が行われるまで出題される。各カテゴリにおけるタスク処理結果から無作為に10,000件の結果データを抽出し、一致率を調査した結果を表2に示す。

表2 各タスクの作業結果における一致率

	3 人一致	2 人一致	不一致	不適切
単語判定カテゴリ (2 択)	71.1%	28.9%	0.0%	0.0%
読み付けカテゴリ	75.6%	16.7%	2.4%	5.3%
品詞カテゴリ	84.3%	2.4%	13.2%	0.1%
アクセントカテゴリ	66.0%	28.5%	3.9%	1.6%

4.4 結果

以上の処理を用いてWebから得られた語彙数を表3に示す。125億文のWebテキストから14万語の語彙を獲得することができた。獲得できた未知語の例としては「Siri」「あっちゃん」「先っちょ」「スゲー」「ドm」「花立山」「えらそう」「やべえええええ」などが挙げられる。このように語彙の収集にクラウドソーシングを用いることによって時間コストと費用を大きく削減することができている。

表3 未知語獲得数

未知語候補抽出数	227,367
未知語獲得数	138,546

5. 今後の課題

本システムにおける課題としてワーカのモチベーションコントロールがある。ワーカのモチベーションコントロールはマイクロタスク型のクラウドソーシングにおけるコスト、速度、精度、すべての点に影響を与える大きな問題である。マイクロタスク型のクラウドソーシングはマッチング型やコンペティション型のクラウドソーシングと比較して1つひとつの作業が小さいため達成感があまり得られない。そのため何らかの目標を与える必要がある。本研究では難易度や高いモチベーションが必要なタスクにおいては報酬の高低でコントロールしているが、報酬によるモチベーションコントロールはマイクロタスク型クラウドソーシングの低コストという利点を損なう可能性が出てきてしまうため、報酬に依存しないモチベーションの向上手法を検討する必要がある。

報酬を用いないワーカのモチベーションのコントロールとしてゲーミフィケーション的なアプローチがある[3]。クラウドソーシングにおけるゲーミフィケーションの適用はタスクの内容に依存するものが多く、本研究におけるシステム側からの適用は難易度が高い。システム側におけるゲーミフィケーションの適用としては以下のような方法が考えられる。

- ワーカ間の競争心を刺激する
 - ランキング設定
 - ライバルワーカの設置
 - 勝敗ルールの設定、報酬への重み付け

- ワーカ同士で協力してタスクの処理や学習をする
 - ーワーカ同士でのグループワークの許可
 - ーアクティブ・ラーニング
- 適切なマイルストーンを設置する
 - ーレベルやアイテムなどのコストを伴わない報酬設定
 - ーレベリングによる優遇措置

特にアクティブ・ラーニングや協調学習に関してはマイクロタスク型クラウドソーシング上でも有効性が予想されるが、これらの適用はワーカからの問合せやクレームが増加することも予測され、適用には慎重な対応が求められる。

参考文献

- 1) 芦川将之, 川村隆浩, 大須賀昭彦: マイクロタスク型クラウドソーシングプラットフォーム環境における精度向上手法の導入と評価, 人工知能学会論文誌, 29(6), pp.503-515 (2014).
- 2) Snow, R., O'Connor, B., Jurafsky, D., Ng, A. Y. : Cheap and Fast But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks, Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, pp.254-263 (2008).
- 3) Ahn, L. and Dabbish, L. : Designing Games with a Purpose, Communications of the ACM, pp.58-67 (2008).
- 4) Ambati, V. et al. : Towards Task Recommendation in Micro-task Markets, Human Computation, pp.1-4 (2011).
- 5) Yuen, M. C., et al. : TaskRec : Probabilistic Matrix Factorization in Task Recommendation in Crowdsourcing systems, International Conference on Neural Information Processing, Springer Berlin Heidelberg, pp.516-525 (2012).
- 6) Vaughan, J. W. : Adaptive Task Assignment for Crowd-sourced Classification, In 30th Intl. Conf. on Machine Learning (ICML) (2013).
- 7) 森 信介, 中田陽介, Graham, N., 河原達也: 点予測による形態素解析, 自然言語処理, Vol.18, No.4, pp.367-381 (2011).

芦川 将之（非会員）masayuki.ashikawa@toshiba.co.jp

1976年生。1999年早稲田大学理工学部情報学科卒業。2001年同大学院理工学研究科修士課程修了。2017年電気通信大学大学院博士課程修了。工学博士。現在、東芝デジタルソリューションズ（株）にてクラウドソーシング、大規模データ処理の研究・開発に従事。

川村 隆浩（正会員）kawamura@ohsuga.is.uec.ac.jp

1994年早稲田大学大学院理工学研究科電気工学専攻修士課程修了。同年、（株）東芝入社。2001～2002年米国カーネギー・メロン大学ロボット工学研究所客員研究員兼任。2003年より電気通信大学大学院情報システム学研究科客員准教授兼任。2007年より大阪大学大学院工学研究科非常勤講師兼任。2015年より科学技術振興機構情報分析室主任調査員。現在に至る。工学博士（早稲田大学）。2012年ISWC 10-Year Award受賞。主として知識抽出、セマンティックWebに関する研究・開発に従事。本会シニア会員。

大須賀 昭彦（正会員）ohsuga@uec.ac.jp

1958年生。1981年上智大学理工学部数学科卒業。同年（株）東芝入社。同社研究開発センター、ソフトウェア技術センター等に所属。1985～1989年（財）新世代コンピュータ技術開発機構（ICOT）出向。2007年より電気通信大学。現在、同大学大学院情報理工学研究科教授。2017年より同大学大学院情報システム学研究科研究科長併任。2012年よ

り国立情報学研究所客員教授兼任。工学博士（早稲田大学）。本会フェロー。ソフトウェア工学，エージェント，人工知能の研究に従事。1986年度および2016年度本会論文賞，2013年度人工知能学会研究会優秀賞，2014年度同学会功労賞受賞。IEEE Computer Society Japan Chapter Chair，人工知能学会理事，日本ソフトウェア科学会理事，同学会監事等を歴任。電子情報通信学会，人工知能学会，日本ソフトウェア科学会，電気学会，IEEE Computer Society 各会員。本会シニア会員。

採録決定：2018年8月7日

編集担当：今原 修一郎（（株）東芝）